# A Gradient-like Variational Bayesian Approach for Unsupervised Extended Emission Map-Making from SPIRE/HERSCHEL DATA

H Ayasso, Thomas Rodet, Alain Abergel

# A GRADIENT-LIKE VARIATIONAL BAYESIAN APPROACH FOR UNSUPERVISED EXTENDED EMISSION MAP-MAKING FROM SPIRE/HERSCHEL DATA

Hacheme Ayasso[1], Thomas Rodet[2] and Alain Abergel[1]

[1]:Institut d'Astrophysique Spatial (UMR8617 CNRS–Univ Paris-Sud)

Paris-Sud, Bat 121, 91405 ORSAY CEDEX, France

[2]: Laboratoire des Signaux et Systèmes, (UMR8506 Univ Paris-Sud–CNRS–SUPELEC)

3 rue Joliot-Curie, 91190 Gif-sur-Yvette, France

**Abstract**

In this work, we study the problem of extended source emission high resolution map-making, which is tackled as an inverse problem. Therefore, an unsupervised Bayesian framework is proposed for estimating sky maps and all the related hyperparameters. For the forward problem, a detailed physical model is introduced to describe different instrument effects like: optical transfer function, pointing process and temperature drifts. Several models for pointing and optical transfer functions are implemented and a Gaussian distribution is attributed to noise model. Since we are interested in extended emission, a Markovian field accounting for four closest neighbors is used as a sky prior.

In our unsupervised approach, we write the joint posterior of the sky and all the hyperparameters (prior correlation and noise parameters) as a function of the likelihood and the different priors. Nevertheless, its expression is complicated and neither the joint maximum a posteriori (JMAP) nor the posterior mean (PM) have an explicit form. Therefore, we propose a new gradient like variational Bayesian approach to tackle the problem of posterior approximation. In order to accelerate the convergence, shaping parameters are updated simultaneously like in a gradient method. We applied our approach for unsupervised map-making of simulated data and real SPIRE/Herschel data. The results show good performance for our method in term of reconstruction quality and hyperparameters estimation and a gain in spatial resolution up to 3 times compared to conventional methods.

## 1 Introduction

Map-making of astronomical imaging instruments has a big importance in astrophysical community since it is the first step for data analysis. Therefore, map-making tools should be able to produce high quality maps that correspond to real sky and have minimum artifacts for the instrument. For example in far-infrared imaging, the space observatory Herschel

[1] was launched in 2009 to help acquiring images of nearby star-forming clouds, galaxies and distance galaxies. Although, Herschel has the largest mirror for a space telescope, its spatial resolution is highly limited by diffraction. Furthermore, its cryogenic design makes it susceptible to temperature drift which reduces the quality of the results. Hence, it became crucial to introduce high resolution map making techniques that reduces the noise while conserving the photometery of the instrument.

Several methods have been used for Herschel map making such as so-called naïve(co-addition) which consist of averaging measurements falling on the same sky pixel, maximum likelihood methods like MADmap [2] and SANEPIC [3] that make a correlated form of noise on measurement error and the Bayesian methods as in [4].

However, these methods suffers from several drawbacks. For example, maps estimated with the first and second methods have limited spatial resolution, since they do not account for instrument optical transfer function in their models. Furthermore, the Bayesian method [4], which accounts for the optical transfer function of the instrument, lacks a proper model for temperature drift and automatic estimation of hyperparameters.

For these reasons, we propose a new unsupervised method which accounts for physical model of the instrument. So, the problem of high resolution map making is tackled as an inverse problem. This is achieved by means of an unsupervised Bayesian framework which permits seamless integration of prior information. We focus in this study on map making of extended sources (dust clouds, $\cdots$). For this, a correlated Gaussian field is used for smooth component modeling. Moreover, a Gaussian distribution with varying mean is used to account for temperature drift. In addition, all the hyperparameters of the model ( noise variance, correlation parameter, offset value ) are estimated jointly with sky map so that the reconstruction method is robust with respect to initial parameters choice by the user.

Nevertheless, the joint posterior has a complex expression and neither the *Joint Maximum A Posteroiri* (JMAP) nor the *Posterior Mean* (PM) estimators have a tractable form. Hence, an approximation is needed to obtain a practical solution. Several methods were proposed in literature such as stochastic sampling by Markov Chain Monte Carlo (MCMC) methods [5] or deterministic like the variational Bayesian approach [6] which approximate the true posterior by a separable free-form distribution. The former method necessitate drawing an important number of random samples from the true posterior in order to calculate an empirical estimator of the PM. This allows good exploration of posterior space, however compared to the latter method, it is more time demanding especially for huge data sets since too many samples are needed to explore the space. Therefore, we opted herein for a new deterministic method [7] based on the variational Bayesian approach since treated data have generally hundred millions samples per observed field. The main contribution of this work is development of an unsupervised superresolution method with detector offset estimation in a gradient like variational Bayesian framework.

In section 2, we introduce our Bayesian approach with reference to the forward and prior models. Then, the new variational Bayesian approach is presented and the expression of approximated posterior is given. Afterward, The method is tested with simulated and real data from Herschel space observatory. Finally, we conclude this work and give some perspectives.

# 2 Bayesian Framework

In this map making problem, we try to restore the sky $\boldsymbol{x}$ given several observations $\boldsymbol{y}$ and instrument model $\mathbf{H}$. More precisely, $\boldsymbol{y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T\}$ is composed of several measurements

$\boldsymbol{y}_i$ covering the sky and mutually shifted by a known translation (figure 1).

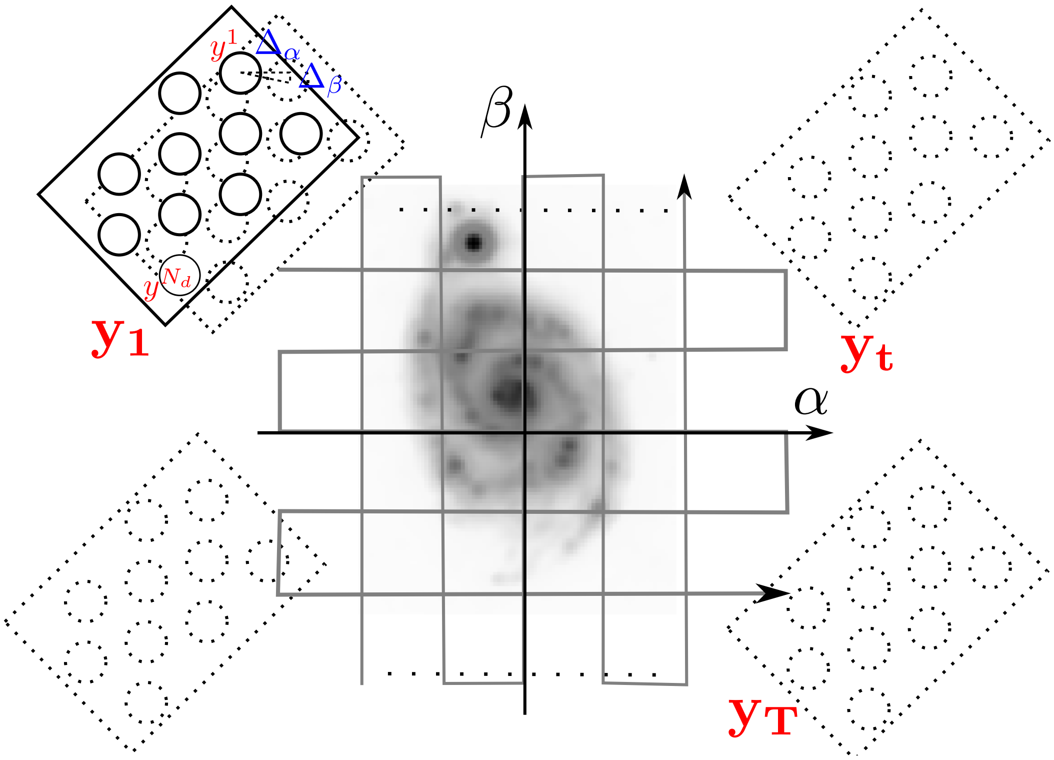


Figure 1: Schematic representation of the scanning process in the telescope.

In case of Herschel, a linear model with additive noise is attributed to the instrument so the forward model reads

$$\boldsymbol{y} = \mathbf{H}\,(\boldsymbol{p} + \boldsymbol{s}) + \boldsymbol{n}. \tag{1}$$

The instrument model $\mathbf{H} = \mathbf{UC}$ is a linear operator containing $\mathbf{U}$ the pointing matrix determining the sky pixels seen by the detector and $\mathbf{C}$ is a convolution matrix accounting for the instrument optical system .

This is an ill-posed problem since the operator $\mathbf{H}$ is ill-conditioned. Therefore, we opt for a Bayesian approach to reduce the dimension of admissible solutions space by introducing a prior distribution based one the sky properties.

The ingredients for the posterior distribution (likelihood and prior) are defined next. Then, the problem of estimation is discussed in the following section.

## 2.1 Likelihood

We assign a white Gaussian distribution for the additive noise with a unknown variance $\rho_n^{-1}$. Furthermore, since Herschel is a Cryogenic instrument, temperature drifts affect measurements by adding a varying offset each detector reading. We model this by an unknown mean $\boldsymbol{o}$ of the noise distribution . Using the forward model (equation 1), likelihood reads,

$$\mathcal{P}(\boldsymbol{y}|\rho_n, \boldsymbol{o}, \mathbf{H}, \boldsymbol{x}) \propto \exp\left(-\frac{\rho_n \left\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x} - \boldsymbol{o}\right\|_2^2}{2}\right) \tag{2}$$

## 2.2 Priors

Prior distribution plays an important role in defining the quality of reconstruction. It is chosen to represent available information over the studied sky. In this work, we are interested in reconstructing skies with extended emission. Therefore, a correlated multivariate Gaussian Markov field, which accounts for smooth variations, was assigned to $\boldsymbol{x}$, so

$$\mathcal{P}(\boldsymbol{s}|\rho_s) \propto \exp\left(-\frac{\rho_x\left(\|\mathbf{D}_\alpha \boldsymbol{x}\|_2^2 + \|\mathbf{D}_\beta \boldsymbol{x}\|_2^2\right)}{2}\right), \tag{3}$$

where $\mathbf{D}_\alpha$ and $\mathbf{D}_\beta$ are finite differences matrices according $\alpha$ and $\beta$ axes respectively, and $\rho_x$ is a parameter determining the degree of correlation in the field which is considered unknown.

For the model hyperparameters $\boldsymbol{\theta} = \{\rho_n, \rho_x, \boldsymbol{o}\}$ conjugate prior were assigned,

$$\begin{aligned} \rho_n &\sim \mathcal{G}(\gamma_n, \phi_n), \\ \rho_x &\sim \mathcal{G}(\gamma_x, \phi_x), \\ \boldsymbol{o} &\sim \mathcal{N}(\boldsymbol{m}_o, \mathbf{V}_o) \end{aligned}$$

Other shaping parameters $(\gamma_n, \phi_n, \gamma_x, \phi_x, \boldsymbol{m}_o, \mathbf{V}_o)$ are fixed to have flat (non-informative) distributions.

All the ingredient are set to have the joint posterior distribution $\mathcal{P}(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. However applying the JMAP or PM yields intractable solution and we need an approximation to obtain one. We discuss in the next section the new variational Bayesian method used for this work.

# 3 Gradient-like variational Bayesian approach

The mutual dependence between different variables make it hard to obtain a tractable solution of the optimization/integration problem needed by the JMAP/PM estimator respectively. Furthermore due to huge space dimensions, simple numerical optimization/integration methods are impossible to apply. Therefore, the variational Bayesian approach, introduced by [8], proposes to approximate the joint posterior $\mathcal{P}(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ by a separable free form distribution $\mathcal{Q}(\boldsymbol{u}) = \prod_i \mathcal{Q}(\boldsymbol{u}_i)$, $\boldsymbol{u} = \{\boldsymbol{x}, \boldsymbol{\theta}\}$ that minimizes the Kullback-Leibler divergence,

$$KL(\mathcal{P}||\mathcal{Q}) = \int \mathcal{Q}(\boldsymbol{u}) \log\left(\frac{\mathcal{P}(\boldsymbol{u}|\boldsymbol{y})}{\mathcal{Q}(\boldsymbol{u})}\right) \mathrm{d}\boldsymbol{u} = \log\left(\mathcal{P}(\boldsymbol{y}|\mathcal{M})\right) + \mathcal{F}(\mathcal{Q}), \tag{4}$$

where $\mathcal{P}(\boldsymbol{y}|\mathcal{M})$ is the model evidence and $\mathcal{F}(\mathcal{Q})$ is the free energy. The functional optimization problem yields an analytical solution for distributions from exponential family, where shaping parameters are mutually dependent and should be updated singly.

A new variational Bayesian approach, proposed by [7], updates the approximating marginals simultaneously. As in a classical gradient method, the shaping parameters are updated simultaneously with an optimal gradient step $\lambda$ which is calculated to maximize free energy $\mathcal{F}(\mathcal{Q})$. So, the approximating marginals have an iterative functional form and their form at iteration $k$ read,

$$\mathcal{Q}_k(\boldsymbol{u}_i) \propto \left(\mathcal{Q}(\boldsymbol{u}_i)\right)^{1-\lambda} \exp\left(\lambda \left\langle \log\left(\mathcal{P}(\boldsymbol{u}, yb)\right)\right\rangle_{\prod\limits_{j\neq i} \mathcal{Q}(\boldsymbol{u}_j)}\right), \tag{5}$$

In this work, we chose a strong separation layout where all the all variables are considered independent. The approximating posterior reads,

$$\mathcal{Q}(\boldsymbol{u}) = \mathcal{Q}(\rho_n)\mathcal{Q}(\rho_x)\prod_i \mathcal{Q}(x_i)\ \prod_j \mathcal{Q}(o_j), \tag{6}$$

and applying eq.5, we obtain approximating marginals from the same family as the priors,

$$\begin{aligned} \check{\mathcal{Q}}(\boldsymbol{x}) &= \mathcal{N}(\check{\boldsymbol{m}}_x, \check{\mathbf{V}}_x), & \check{\mathcal{Q}}(\rho_n) &= \mathcal{G}(\check{\gamma}_n, \check{\phi}_n), \\ \check{\mathcal{Q}}(\rho_x) &= \mathcal{G}(\check{\gamma}_x, \check{\phi}_x), & \check{\mathcal{Q}}(\boldsymbol{o}) &= \mathcal{N}(\check{\boldsymbol{m}}_o, \check{\mathbf{V}}_o), \end{aligned}$$

Furthermore, $\boldsymbol{x}$ components are updated simultaneously with a gradient step $\lambda_x$ and separately from $\rho_n$, $\rho_x$, and $\boldsymbol{o}$ whose step values are fixed to 1 to accelerate the convergence. The shaping parameters are given as,

$$\check{\mathbf{V}}_x^k = \left[(1-\lambda_x)(\check{\mathbf{V}}_x^{k-1})^{-1} + \lambda_x \mathbf{Diag}\left(\bar{\rho}_n \mathbf{H}^t\mathbf{H} + \bar{\rho}_x\left(\mathbf{D}_\alpha^t\mathbf{D}_\alpha + \mathbf{D}_\beta^t\mathbf{D}_\beta\right)\right)\right]^{-1} \tag{7}$$

$$\check{\boldsymbol{m}}_x^k = \check{\boldsymbol{m}}_x^{k-1} + \lambda_x \check{\mathbf{V}}_x^k\Big(\bar{\rho}_n\mathbf{H}^t(\boldsymbol{y} - \check{\boldsymbol{m}}_o - \mathbf{H}\check{\boldsymbol{m}}_x - \bar{\rho}_x(\mathbf{D}_\alpha^t\mathbf{D}_\alpha + \mathbf{D}_\beta^t\mathbf{D}_\beta)\check{\boldsymbol{m}}_x\Big) \tag{8}$$

$$\check{\phi}_n^k = \left(\phi_n + \frac{N_y}{2}\right), \tag{9}$$

$$\check{\gamma}_n^k = \left[\frac{2\gamma_n^{-1} + \left(\|\boldsymbol{y} - \check{\boldsymbol{m}}_o - \mathbf{H}\check{\boldsymbol{m}}_x\|_2^2 + \left\|\check{\mathbf{V}}_o\right\|_1 + \mathbf{H}^t\mathbf{H} : \check{\mathbf{V}}_x\right)}{2}\right]^{-1} \tag{10}$$

$$\check{\gamma}_x^k = \left(\gamma_x^{-1} + \frac{\|\mathbf{D}_\alpha\check{\boldsymbol{m}}_x\|_2^2 + \|\mathbf{D}_\beta\check{\boldsymbol{m}}_x\|_2^2 + (\mathbf{D}_\alpha^t\mathbf{D}_\alpha + \mathbf{D}_\beta^t\mathbf{D}_\beta) : \check{\mathbf{V}}_x}{2}\right)^{-1}, \tag{11}$$

$$\check{\mathbf{V}}_o^k = (\rho_o + \bar{\rho}_n N_d)^{-1}, \tag{12}$$

$$\check{\boldsymbol{m}}_o^k = \check{\mathbf{V}}_o^k\left(\boldsymbol{m}_o\rho_o + \bar{\rho}_n\sum_{i\in\Omega}\boldsymbol{y}_i - \check{\boldsymbol{y}}_i\right), \tag{13}$$

with

$$\begin{aligned} \bar{\rho}_n &= \check{\gamma}_n\check{\phi}_n, & \bar{\rho}_x &= \check{\gamma}_x\check{\phi}_x, \\ N_y &= N_d \times N_T = Dim(\boldsymbol{y}), & N_s &= Dim(\boldsymbol{s}), \\ \|\mathbf{A}\|_1 &= \sum_{i,j} a_{ij}, & \mathbf{A} : \mathbf{B} &= \sum_{i,j} a_{ij}b_{ij}, \\ \check{\boldsymbol{y}} &= \mathbf{H}\check{\boldsymbol{m}}_x, & \mathbf{Diag}\left(\mathbf{A}\right) &= \mathbf{I}\circ\mathbf{A}, \end{aligned}$$

$\mathbf{C} = \mathbf{A}\circ\mathbf{B}$ is the Hadamard product (element-wise $c_{i,j} = a_{i,j}b_{i,j}$), and $\Omega$ is the set of observation for which the detectors offset is considered fixed. In the following, we present several tests of the method on simulated and real data.

# 4 Results

The proposed method was tested using simulated and real data from the space observatory Herschel. For simulation, two simulated fields were used. The first is an extended field generated using a sample from correlated Gaussian field ($\rho_s = 200$) and a white Gaussian noise

($\rho_n = 600$) is added to the forward model output. This test is useful to evaluate the quality of construction when the fields correspond to our model but also the quality of hyperparameter estimation. Figure (2) shows restoration results for our method in accordance with the original map and a significant enhancement compared to *Coadd* method ($\boldsymbol{x}_c = \frac{\mathbf{U}^t \boldsymbol{y}}{\mathbf{U}^t \mathbb{1}}$) used in the official data processing product [9]. The relative error between the real and the coadded map is $\sqrt{\frac{\|\boldsymbol{x}_{Coadd} - \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}} = 8\%$ meanwhile with our method achieves a relative error of 4% only.

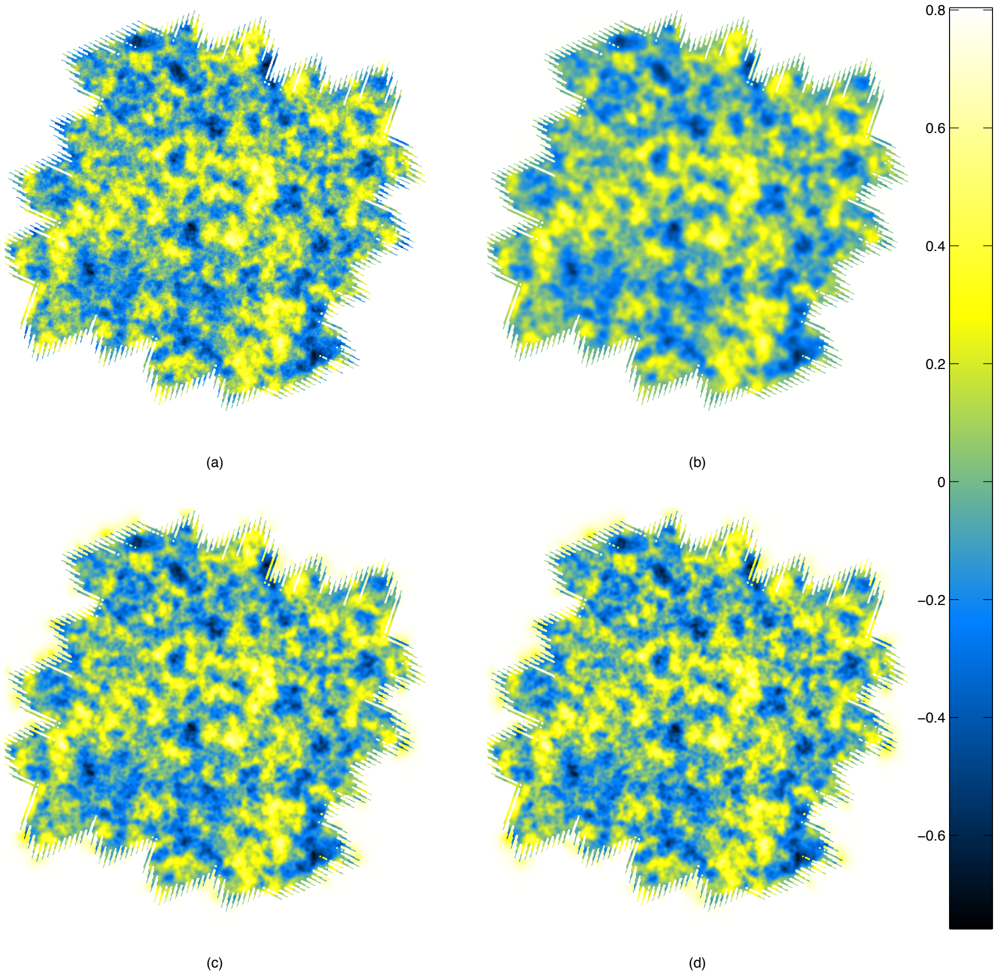


Figure 2: Simulation results. (a) Original map $\boldsymbol{x}$, (b) coadded $\boldsymbol{x}_{Coadd}$, (c) our method $\hat{\boldsymbol{x}}$ unsupervised, (d) our method $\hat{\boldsymbol{x}}$ supervised

Furthermore, the method is able to estimate the hyperparameters with a small error (figure (3)). At the convergence, the parameters reach values close to the true ones as expected since it is hardly attainable because of the uncertainty on $\boldsymbol{x}$ [1]. Moreover in comparison with a supervised reconstruction using the true hyperparameters values (figure (2.d)), the relative error is only 2%

The second simulated field uses a high resolution optical image of Messier galaxy to test the capacity of our method to restore realistic fields. After applying the model to test image, a white Gaussian noise is added with variance $\rho_n^{-1} = 0.14$ and a random offset vector to simulate the temperature drift in the detectors.

A general comparison (figure 4) of reconstruction results confirm the capacity of our method to restore high frequencies. Moreover, taking a closer look (figure 5), we see a strip

[1] The uncertainty of variables is presented in terms $\left\|\check{\mathbf{V}}_o\right\|_1, \mathbf{H}^t\mathbf{H} : \check{\mathbf{V}}_x$ and $(\mathbf{D}_\alpha^t \mathbf{D}_\alpha + \mathbf{D}_\beta^t \mathbf{D}_\beta) : \check{\mathbf{V}}_s x$

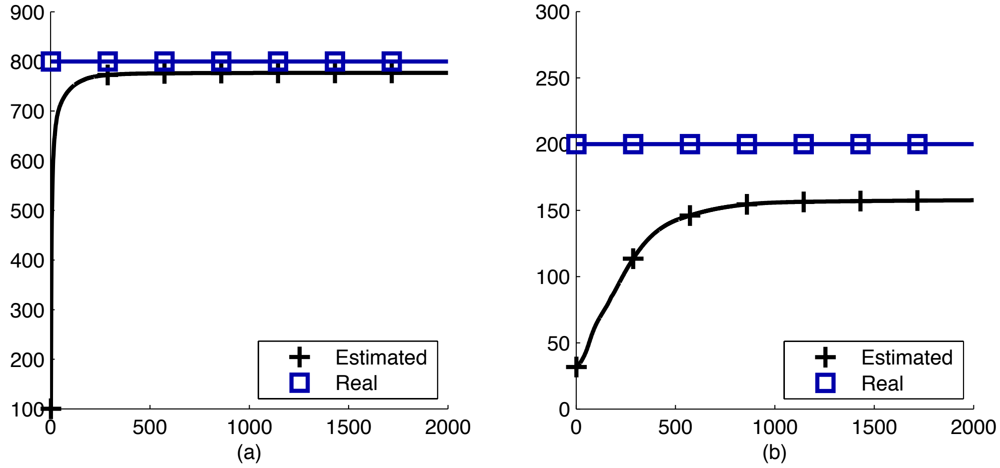


Figure 3: Hyperparameters evolution with iterations. (a) noise precision $\rho_n$, (b) correlation parameter $\rho_x$

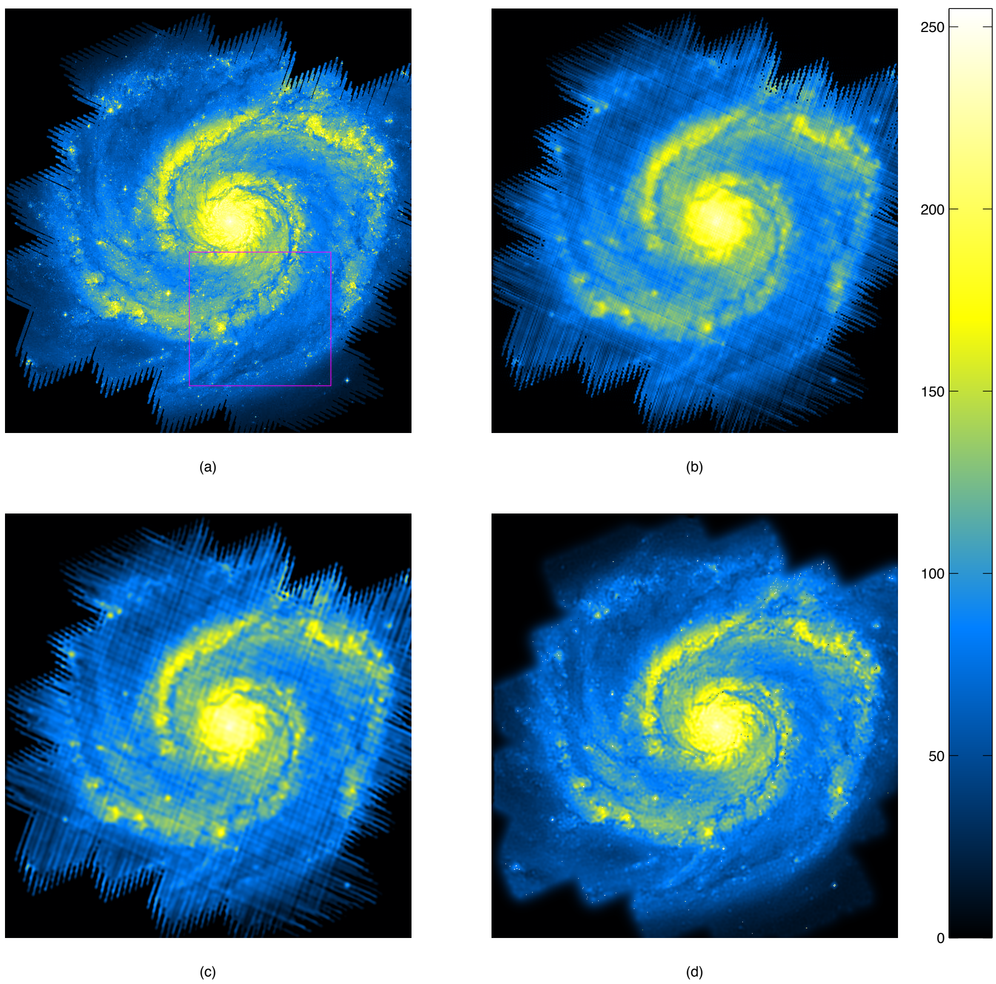


Figure 4: Comparison between different map making methods for Messier galaxy. a) True sky, b) Coaddition map, c) Our method without offset estimation, d) our method with offset estimation .

effect present for methods not accounting for offsets since different detectors have different values when observing the same sky region. Nevertheless, this effect was corrected by the proposed method.

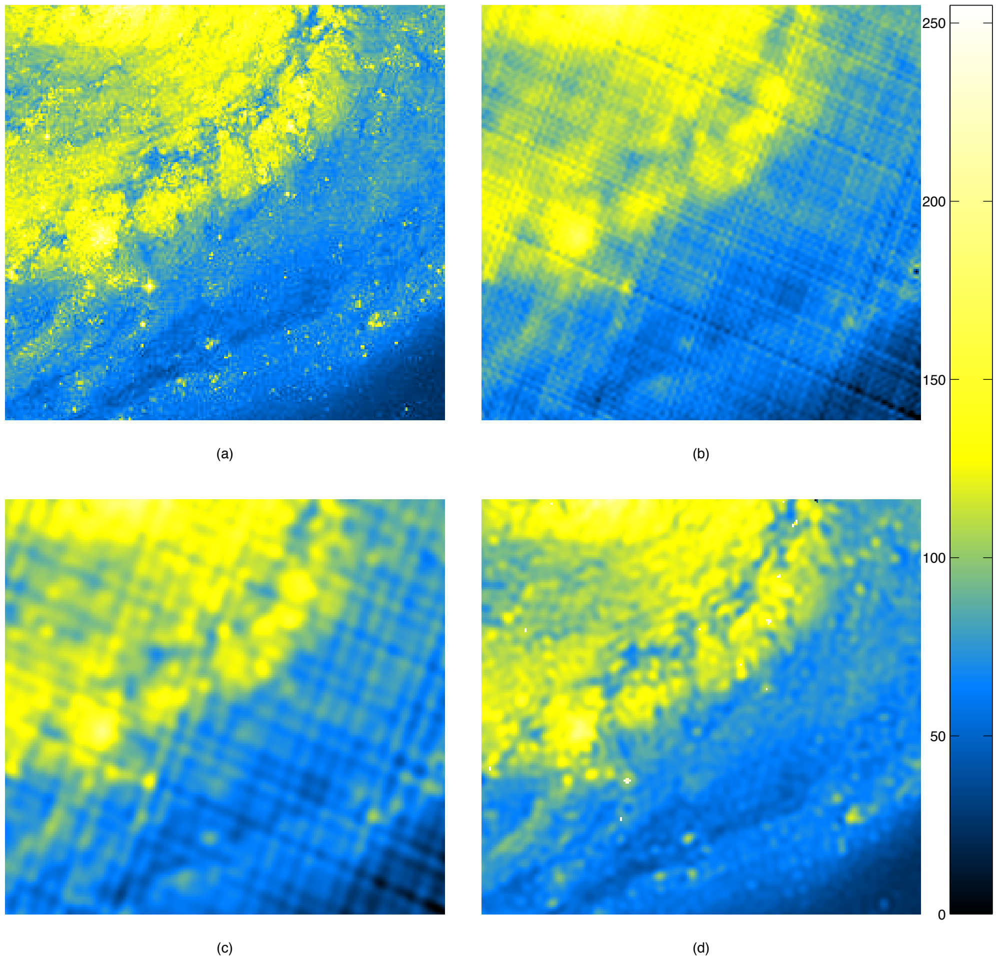


Figure 5: Enlarged bottom right corner (magenta rectangle in figure 4.a) for Messier galaxy to show the offset effect . a) True sky, b) Coaddition map, c) Our method without offset estimation, d) our method with offset estimation $\hat{\boldsymbol{x}}$

Moreover, we can compare the estimated offset values with real one. Figure (6) shows a comparison between the real and estimated offset values and the relative error of estimation. With maximum relative of 1.8%, our method is able to estimate the detector offset with small error.

The last test is applied on a real observation data from SPIRE/Herschel space observatory of the horsehead nebula. We compare mapmaking results for *coaddition* method and our methods in PSW band ($250\mu m$) and use Coadd maps from PACS instrument ($160\mu m$ and $70\mu m$ ) as a reference since they have higher spatial resolution (figure (7)). Our mapmaking reconstruction have clearly more spatial structures than the coadd one. Moreover in comparison with PACS maps, we find a high correlation in the structures with our method which indicate that these structures are real.

# 5 Conclusions

A new superresolution method has been presented for SPIRE/Herschel mapmaking. We adopted an unsupervised Bayesian framework with prior modeling for extended emission

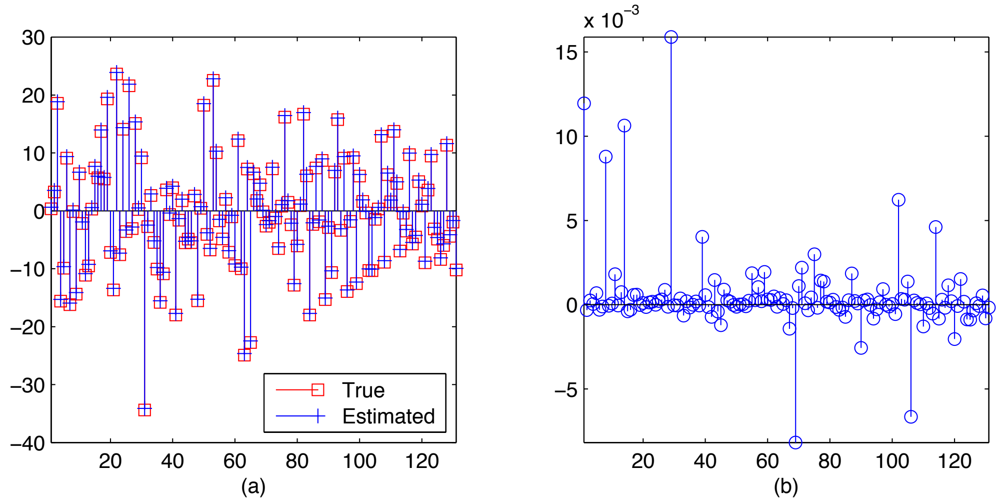


Figure 6: Offset estimation study for Messier galaxy. a) Real $\boldsymbol{o}$ and estimated $\check{\boldsymbol{m}}_o$ offsets comparison, b) relative error.

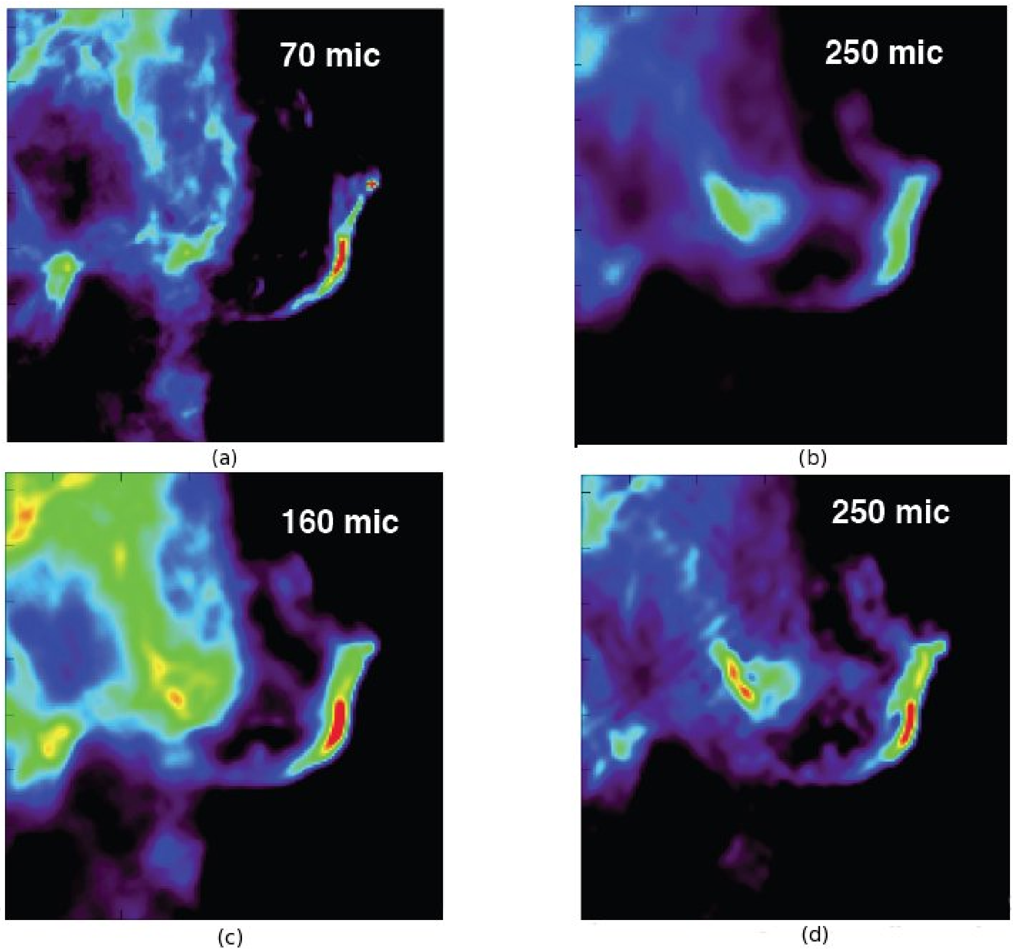


Figure 7: Reconstruction results for Horsehead nebula (real data Herschel). (a) Coadd map form instrument PACS ($70\mu m$) , (b) Coadd map from instrument SPIRE $250\mu m$, (c) Coadd map PACS ($160\mu m$), (d) our method map from instrument SPIRE $250\mu m$

and detectors offsets. For the estimation, a new gradient like variational Bayesian method was used. the performance of the method was tested using several datasets of simulated and real data and it showed good improvement in spatial resolution. Moreover, the automatic estimation performance allows to have a robust method to hyperparameter choice.

Nevertheless, this method contain few drawbacks which needs a special attention. For example, the unconvexiy of the variational Bayesian approach, makes it important to pay special attention initialization to avoid local solution. Furthermore for some fields detectors drift varies in important level across the same scanning leg, which necessitate a more adapted model as the case of [3]. Moreover due to the high non-stationnarity of certain fields (presence of mixture of slowly and highly varying structures ) may be better modeled by a non-homogenous prior.

# References


[1] G. L. et al Pilbratt. Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy. *Astronomy & Astrophysics*, 518:L1, July 2010.

[2] C. M. Cantalupo, J. D. Borrill, A. H. Jaffe, T. S. Kisner, and Stompor. Madmap: A massively parallel maximum likelihood cosmic microwave background map-maker. *The Astrophysical Journal Supplement*, pages 187–212, 2010.

[3] G. Patanchon, P. A. R. Ade, J. J. Bock, E. L. Chapin, M. J. Devlin, S. Dicker, M. Griffin, J. O. Gundersen, M. Halpern, P. C. Hargrave, D. H. Hughes, J. Klein, G. Marsden, P. G. Martin, P. Mauskopf, C. B. Netterfield, L. Olmi, E. Pascale, M. Rex, D. Scott, C. Semisch, M. D. P. Truch, C. Tucker, G. S. Tucker, M. P. Viero, and D. V. Wiebe. Sanepic: A map-making method for timestream data from large arrays. *The Astrophysical Journal*, 681:708, 2008.

[4] F. et al Orieux. Super-resolution in map-making based on a physical instrument model and regularized inversion. application to spire/herschel. *Astronomy & Astrophysics*, AA/2011/16817:16, 2011.

[5] C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Verlag, 2004.

[6] V. Smídl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer Verlag, Berlin, 2006.

[7] A. Fraysse and T. Rodet. A gradient-like variational bayesian algorithm. In *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pages 605 –608, june 2011.

[8] G. E. Hinton and D. van Camp. Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the sixth Annual Conference on Computational Learning Theory*, pages 5–13. ACM New York, NY, USA, 1993.

[9] ESA. *Herschel Interactive Processing Environment (HIPE)*,.